INF 2178 Technical Assignment 4

Jinhang Luo (1005124777)

Faculty of Information
University of Toronto
INF2178: Experimental Design for Data Science
Prof.Shion Guha

April 5st, 2024

Introduction

1.1 Background

This dataset, sourced from a longitudinal MRI study that compares demented and nondemented patients, such as tracking changes in brain volume and cognitive function over different visit times. It offers valuable insights into dementia's progression, which is crucial for improving diagnostics, treatments, and patient outcomes through early interventions.

1.2 Dataset Description

The dataset contains 294 records and numerous attributes (Boysen, 2017). However, our main focus is exploring the interaction between variables such as Normalized whole-brain volume (nWBV), Mini-Mental State Examination (MMSE), patient group (Demented/Nondemented), and visit times.

1.3 Research question

In analyzing the dataset, we aimed to explore how dementia status and visit times affect changes in brain volume and cognitive function. Consequently, two research questions have been formed.

1. How does MMSE differ between groups (Demented/Nondemented) across different visit times?
2. How does the interaction effect between groups (Demented/Nondemented) and different visit times affect nWBV?

1.4 Data Cleaning

We tidied up the dataset by removing unnecessary columns and filling missing values with mean.

Exploratory Data Analysis

2.1 Descriptive Statistics

First, we performed a descriptive analysis that included the dataset's sample size, mean, and standard deviation for certain attributes. As shown in **Table 1**, the sample size of the data is 294. Moreover, the variability in means and standard deviations among various variables suggests that the dataset is robust enough to provide meaningful statistical conclusions.

Table 1. Descriptive Statistics

Variables	N	Mean	SD
Age	294	76.412	7.607
Mini-Mental State Examination	294	27.259	3.408
Normalized whole-brain volume	294	0.731	0.037
Visit Times	294	1.490	0.501

Source: MRI and Alzheimers - Longitudinal Study on MRI From Kaggle

2.2 Distribution Visualizations

As shown in **Figure 1**, these histograms visually illustrate overall age is skewed, and the majority of people have a high concentration of MMSE, suggesting that cognitive functioning is generally stable. nWBV has a consistent distribution centered at 0.75. In the box plots, the age range of the 'converted' is slightly older, and the median MMSE scores are higher in the 'non-demented' subjects. Similarly, the nWBV is also higher.

These plots demonstrate the potential interaction of cognitive function and brain structure in the different dementia status groups and visit times, respectively.

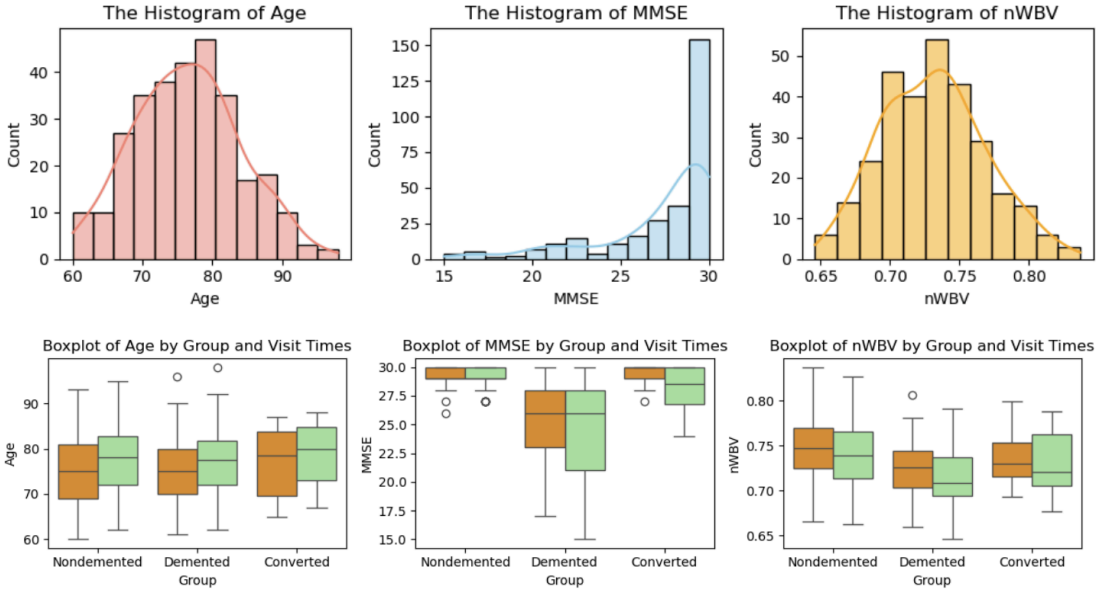

Figure 1: The Histogram and Boxplot Plots of all Variables

Quantitative Analysis (Mixed-Effects ANOVA)

3.1 Output Plot

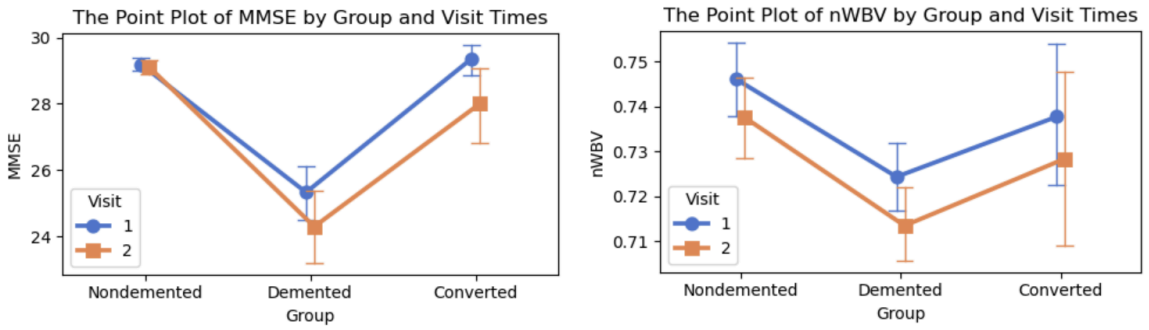

Figure 2: The Point Plots of MMSE/nWBV by Group and Visit Times

As shown in **Figure 2**, the point plots demonstrate clear trends in MMSE and nWBV across different dementia status groups and visit times. For MMSE, all groups showed a decrease at the second visit, with the dementia group showing the most significant reduction. For nWBV, both the demented and converted groups showed a decrease in Normalized whole-brain volume at the second visit, suggesting progressive brain atrophy over time.

3.2 Perform Mixed-Effects ANOVA

Table 2. ANOVA Summary (MMSE as Outcome Variable)

Source	SS	DF1	DF2	MS	F	p–unc	np2	eps
Group	1322.017	2	141	661.009	56.100	<0.001	0.443	NaN
Visit Times	21.528	1	141	21.528	8.525	0.004	0.057	1
Interaction	16.204	2	141	8.102	3.208	0.043	0.044	NaN

Source: MRI and Alzheimers - Longitudinal Study on MRI From Kaggle

Table 3. ANOVA Summary (nWBV as Outcome Variable)

Source	SS	DF1	DF2	MS	F	p–unc	np2	eps
Group	0.034	2	141	0.017	6.712	0.002	0.087	NaN
Visit Times	0.007	1	141	0.007	94.251	<0.001	0.401	1
Interaction	<0.001	2	141	<0.001	1.534	0.219	0.021	NaN

Source: MRI and Alzheimers - Longitudinal Study on MRI From Kaggle

As seen in **Table 2** and **Table 3**, the p-values for both group and visit times on MMSE are less than $\alpha = 0.05$, suggesting strong evidence to reject the null hypothesis. This also suggests a statistically significant difference in MMSE between groups and across visit times. Similarly, for nWBV, group and visit times indicate p-values less than $\alpha = 0.05$, again showing significant effects. However, the interaction effect between group and visit times on nWBV is insignificant, with a p-value greater than $\alpha = 0.05$, suggesting no significant interaction effect on Normalized whole-brain volume changes over time.

3.3 Assumption Checks

a. Mauchly's test of sphericity

Mauchly's test of sphericity resulted in a value of 1 ($p > 0.05$) for both models, suggesting that the assumption of sphericity is valid, thus ensuring the reliability of the ANOVA results.

b. *Normality of Residuals*

As shown in **Table 4,** the Shapiro-Wilk test for Model 1 indicates that none of the variables meets the normal distribution with a p-value lower than 0.05. Conversely, Model 2 meets the normal distribution with a W-value higher than 0.95 and a p-value greater than 0.05, validating the assumption of normality. Considering the robustness of the ANOVA to certain violations of its assumptions, we decided to continue with the analysis and will consider other methods for specific explanations in the final limitations section.

Table 4. The Shapiro-Wilk Test. (Model1:MMSE, Model2: nWBV)

	Nondemented (Model1|Model2)	Demented (Model1|Model2)	Converted (Model1|Model2)
Test statistics (W)	0.810|0.989	0.928|0.990	0.771|0.958
P-Value	2.636e-12|0.317	4.487e-06|0.540	5.868e-05|0.358

Source: MRI and Alzheimers - Longitudinal Study on MRI From Kaggle

c. *Homogeneity of Variances*

As shown in **Table 5**, Levene's test for Model 1 shows a significant deviation from homogeneity of variance, with a test statistic (W) of 64.643 and a p-value far below the 0.05 threshold. In contrast, Model 2 meets the homogeneity assumption suggested by Levene's test statistic (W) of 0.949 and a p-value above 0.05, indicating equal variances across groups. Considering the robustness of the ANOVA to certain violations of its assumptions, we decided to continue with the analysis and will consider other methods for specific explanations in the final limitations section.

Table 5. The Levene's Test. (Model1:MMSE, Model2: nWBV)

	Test statistics (W)	P-Value
Value (Model 1)	64.643	5.897e-24
Value (Model 2)	0.949	0.38841

Source: MRI and Alzheimers - Longitudinal Study on MRI From Kaggle

3.4 Perform Statistical Power for T-Tests

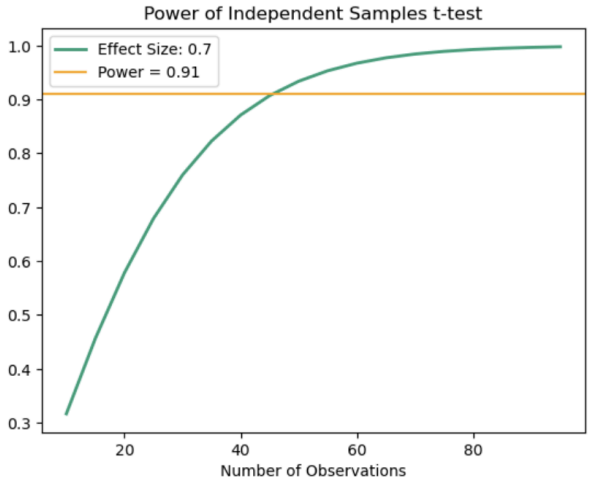

Figure 3: The Power of T-Test

A sample size of 46 was computed to meet the power of 0.91 (alpha value of 0.05 and effect size of 0.7). As shown **in Figure 3**, the power curve also indicates that a sample size of 46 is adequate to attain a power of 0.91, considering an alpha of 0.05 and an effect size of 0.7.

Limitation

1. The findings from this dataset may be limited to other cities or regions.
2. The Shapiro-Wilk and Levene's tests suggest that Model 1 fails to meet the assumptions of normality and homogeneity of variance. While ANOVA is robust to some deviations from this assumption, the extent of the deviation here may have influenced the investigation's results. Alternative methods, such as Welch's ANOVA or Generalized Least Squares, may be considered to address this issue in later stages.

Conclusion

In summary, our investigation demonstrates the impact of the complex interaction between dementia status and visit times on changes in cognitive function and brain volume. The data indicate that Mini-Mental State Examination (MMSE) declines over time, particularly in the dementia group, while Normalized whole-brain volume (nWBV) declines consistently across all groups. This highlights the subtle impact of dementia progression on mental and neurological health, providing important information for clinical approaches aimed at early detection and intervention. These findings supply valuable information for healthcare professionals and researchers on the management of dementia and contribute to research on how to mitigate and prevent the effects of dementia.

Reference List

Boysen, J. (2017, August 16). *MRI and Alzheimers*. Kaggle. https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers?resource=download